

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (i) APPLICANT:  
(A) NAME: Kossmann, Jens  
Buttcher, Volker  
Welsh, Thomas
- (ii) TITLE OF INVENTION: DNA SEQUENCES ENCODING ENZYMES CAPABLE  
OF FACILITATING THE SYNTHESIS OF LINEAR  
ALPHA-1,4 GLUCANS IN PLANTS, FUNGI AND  
MICROORGANISMS
- (iii) NUMBER OF SEQUENCES: 4
- (iv) CORRESPONDENCE ADDRESS:  
(A) ADDRESSEE: James F. Haley, Jr., c/o Fish & Neave  
(B) STREET: 1251 Avenue of the Americas  
(C) CITY: New York  
(D) STATE: New York  
(E) COUNTRY: United States of America  
(F) ZIP: 10020
- (v) COMPUTER READABLE FORM:  
(A) MEDIUM TYPE: Floppy disk  
(B) COMPUTER: IBM PC compatible  
(C) OPERATING SYSTEM: PC-DOS/MS-DOS  
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
- (vi) PRIOR APPLICATION DATA:  
(A) APPLICATION NUMBER: DE P 44 17 879.4  
(B) FILING DATE: 18-MAY-1994
- (vi) PRIOR APPLICATION DATA:  
(A) APPLICATION NUMBER: DE P 44 47 388.5  
(B) FILING DATE: 22-DEC-1994
- (vii) ATTORNEY/AGENT INFORMATION:  
(A) NAME: James F. Haley, Jr., Esq.  
(B) REGISTRATION NUMBER: 27,794  
(C) DOCKET NUMBER: GFB-1
- (viii) TELECOMMUNICATION INFORMATION:  
(A) TELEPHONE: (212) 596-9000  
(B) TELEFAX: (212) 596-9090

## (2) INFORMATION FOR SEQ ID NO: 1:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2914 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *Neisseria polysaccharea*

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: genomic library in pBluescriptII SK

(B) CLONE: pNB2

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 957..2867

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GAGTTTTGCG TTCCCGAACC GAACGTGATG CTTGAGCCGA ACACCTGTCC GGCAAGGCGG	60
CTGACCGCCC CCTTTTGCCC CATCGACATC GTAACAATCG GTTTGGTGGC AAGCTCTTTC	120
GCTTTGAGCG TGGCAGAAAG CAAAGTCAGC ACGTCTTCCG CGCTTTGCGG CATCACCGCA	180
ATTTTGCAGA TGTCCGCGCC GCAGTCCTCC ATCTGTTTCA GACGGCATAAC GATTTCTTCT	240
TGCGGCGGCG TGCGGTGAAA CTCATGATTG CAGAGCAGGG CGGCGATGCC GTTTTTTTGA	300
GCATGCGCCA CGGCGCGCCG GACGGCGGTT TCGCCGAAA AAAGCTCGAT ATCGATAATG	360
TCGGGCAGGC GGCTTTCAAT CAGCGAGTCG AGCAGTTCAA AATAATAATC GTCCGAACAC	420
GGGAACGAGC CGCCTTCGCC ATGCCGTCTG AACGTAAACA GCAGCGGCTT GTCGGGCAGC	480
GCGTCGCGGA CGGTCTGCGT GTGGCGCAAT ACTTCGCCGA TGCTGCCCCG GCATTCCAAA	540
AAATCGGCGC GGAACCTCGAC GATATCGAAG GGCAGGTTTT TGATTTGGTC AAGTACGGCG	600
GAAAGTACGG CGGCATCGCG GGCGACAAGC GGCACGGCGA TTTTGGTGCG TCCGCTTCCG	660
ATAACGGTGT TTTTGACGGT CAGGCTGGTG TGCATGGCGG TTGTTGCGGC TGAAAGGAAC	720
GGTAAAGACG CAATTATAGC AAAGGCACAG GCAATGTTTC AGACGGCATT TCTGTGCGGC	780
CGGCTTGATA TGAATCAAGC AGCATCCGCA TATCGGAATG CAGACTTGGC ACAAGCCCTG	840
TCTTTTCTAG TCAGTCCGCA GTTCTTGCA TATGATTGCA CGACACGCCC TACACGGCAT	900
TTGCAGGATA CGGCGGCAGA CCGCCGGTCG GAAACTTCAG AATCGGAGCA GGCATC	956
ATG TTG ACC CCC ACG CAG CAA GTC GGT TTG ATT TTA CAG TAC CTC AAA	1004
Met Leu Thr Pro Thr Gln Gln Val Gly Leu Ile Leu Gln Tyr Leu Lys	
1 5 10 15	
ACA CGC ATC TTG GAC ATC TAC ACG CCC GAA CAG CGC GCC GGC ATC GAA	1052
Thr Arg Ile Leu Asp Ile Tyr Thr Pro Glu Gln Arg Ala Gly Ile Glu	
20 25 30	
AAA TCC GAA GAC TGG CGG CAG TTT TCG CGC CGC ATG GAT ACG CAT TTC	1100

Lys	Ser	Glu	Asp	Trp	Arg	Gln	Phe	Ser	Arg	Arg	Met	Asp	Thr	His	Phe	
		35					40					45				
CCC	AAA	CTG	ATG	AAC	GAA	CTC	GAC	AGC	GTG	TAC	GGC	AAC	AAC	GAA	GCC	1148
Pro	Lys	Leu	Met	Asn	Glu	Leu	Asp	Ser	Val	Tyr	Gly	Asn	Asn	Glu	Ala	
	50					55					60					
CTG	CTG	CCT	ATG	CTG	GAA	ATG	CTG	CTG	GCG	CAG	GCA	TGG	CAA	AGC	TAT	1196
Leu	Leu	Pro	Met	Leu	Glu	Met	Leu	Leu	Ala	Gln	Ala	Trp	Gln	Ser	Tyr	
	65					70				75					80	
TCC	CAA	CGC	AAC	TCA	TCC	TTA	AAA	GAT	ATC	GAT	ATC	GCG	CGC	GAA	AAC	1244
Ser	Gln	Arg	Asn	Ser	Ser	Leu	Lys	Asp	Ile	Asp	Ile	Ala	Arg	Glu	Asn	
				85					90					95		
AAC	CCC	GAT	TGG	ATT	TTG	TCC	AAC	AAA	CAA	GTC	GGC	GGC	GTG	TGC	TAC	1292
Asn	Pro	Asp	Trp	Ile	Leu	Ser	Asn	Lys	Gln	Val	Gly	Gly	Val	Cys	Tyr	
			100					105					110			
GTT	GAT	TTG	TTT	GCC	GGC	GAT	TTG	AAG	GGC	TTG	AAA	GAT	AAA	ATT	CCT	1340
Val	Asp	Leu	Phe	Ala	Gly	Asp	Leu	Lys	Gly	Leu	Lys	Asp	Lys	Ile	Pro	
		115					120					125				
TAT	TTT	CAA	GAG	CTT	GGT	TTG	ACT	TAT	CTG	CAC	CTG	ATG	CCG	CTG	TTT	1388
Tyr	Phe	Gln	Glu	Leu	Gly	Leu	Thr	Tyr	Leu	His	Leu	Met	Pro	Leu	Phe	
	130					135					140					
AAA	TGC	CCT	GAA	GGC	AAA	AGC	GAC	GGC	GGC	TAT	GCG	GTC	AGC	AGC	TAC	1436
Lys	Cys	Pro	Glu	Gly	Lys	Ser	Asp	Gly	Gly	Tyr	Ala	Val	Ser	Ser	Tyr	
	145				150					155					160	
CGC	GAT	GTC	AAT	CCG	GCA	CTG	GGC	ACA	ATA	GGC	GAC	TTG	CGC	GAA	GTC	1484
Arg	Asp	Val	Asn	Pro	Ala	Leu	Gly	Thr	Ile	Gly	Asp	Leu	Arg	Glu	Val	
				165					170					175		
ATT	GCT	GCG	CTG	CAC	GAA	GCC	GGC	ATT	TCC	GCC	GTC	GTC	GAT	TTT	ATC	1532
Ile	Ala	Ala	Leu	His	Glu	Ala	Gly	Ile	Ser	Ala	Val	Val	Asp	Phe	Ile	
			180					185					190			
TTC	AAC	CAC	ACC	TCC	AAC	GAA	CAC	GAA	TGG	GCG	CAA	CGC	TGC	GCC	GCC	1580
Phe	Asn	His	Thr	Ser	Asn	Glu	His	Glu	Trp	Ala	Gln	Arg	Cys	Ala	Ala	
		195					200					205				
GGC	GAC	CCG	CTT	TTC	GAC	AAT	TTC	TAC	TAT	ATT	TTC	CCC	GAC	CGC	CGG	1628
Gly	Asp	Pro	Leu	Phe	Asp	Asn	Phe	Tyr	Tyr	Ile	Phe	Pro	Asp	Arg	Arg	
	210					215					220					
ATG	CCC	GAC	CAA	TAC	GAC	CGC	ACC	CTG	CGC	GAA	ATC	TTC	CCC	GAC	CAG	1676
Met	Pro	Asp	Gln	Tyr	Asp	Arg	Thr	Leu	Arg	Glu	Ile	Phe	Pro	Asp	Gln	
	225				230					235					240	
CAC	CCG	GGC	GGC	TTC	TCG	CAA	CTG	GAA	GAC	GGA	CGC	TGG	GTG	TGG	ACG	1724
His	Pro	Gly	Gly	Phe	Ser	Gln	Leu	Glu	Asp	Gly	Arg	Trp	Val			

TTC Phe	CGC Arg	GCA Ala 275	ATG Met	GCG Ala	GGC Gly	GAA Glu 280	ATG Met	CTG Leu	TTC Phe	CTT Leu	GCC Ala	AAC Asn 285	TTG Leu	GGC Gly	GTT Val	1820
GAC Asp	ATC Ile 290	CTG Leu	CGT Arg	ATG Met	GAT Asp	GCG Ala 295	GTT Val	GCC Ala	TTT Phe	ATT Ile	TGG Trp 300	AAA Lys	CAA Gln	ATG Met	GGG Gly	1868
ACA Thr 305	AGC Ser	TGC Cys	GAA Glu	AAC Asn 310	CTG Leu	CCG Pro	CAG Gln	GCG Ala	CAC His	GCC Ala 315	CTC Leu	ATC Ile	CGC Arg	GCG Ala	TTC Phe 320	1916
AAT Asn	GCC Ala	GTT Val	ATG Met	CGT Arg 325	ATT Ile	GCC Ala	GCG Ala	CCC Pro	GCC Ala 330	GTG Val	TTC Phe	TTC Phe	AAA Lys	TCC Ser 335	GAA Glu	1964
GCC Ala	ATC Ile	GTC Val	CAC His 340	CCC Pro	GAC Asp	CAA Gln	GTC Val 345	GTC Val	CAA Gln	TAC Tyr	ATC Ile	GGG Gly	CAG Gln 350	GAC Asp	GAA Glu	2012
TGC Cys	CAA Gln 355	ATC Ile	GGT Gly	TAC Tyr	AAC Asn	CCC Pro	CTG Leu 360	CAA Gln	ATG Met	GCA Ala	TTG Leu 365	TTG Leu	TGG Trp	AAC Asn	ACC Thr	2060
CTT Leu 370	GCC Ala	ACG Thr	CGC Arg	GAA Glu	GTC Val 375	AAC Asn	CTG Leu	CTC Leu	CAT His	CAG Gln 380	GCG Ala	CTG Leu	ACC Thr	TAC Tyr	CGC Arg	2108
CAC His 385	AAC Asn	CTG Leu	CCC Pro	GAG Glu 390	CAT His	ACC Thr	GCC Ala	TGG Trp	GTC Val 395	AAC Asn	TAC Tyr	GTC Val	CGC Arg	AGC Ser	CAC His 400	2156
GAC Asp	GAC Asp	ATC Ile	GGC Gly 405	TGG Trp	ACG Thr	TTT Phe	GCC Ala	GAT Asp	GAA Glu 410	GAC Asp	GCG Ala	GCA Ala	TAT Tyr	CTG Leu 415	GGC Gly	2204
ATA Ile	AGC Ser	GGC Gly 420	TAC Tyr	GAC Asp	CAC His	CGC Arg	CAA Gln 425	TTC Phe	CTC Leu	AAC Asn	CGC Arg	TTC Phe 430	TTC Phe	GTC Val	AAC Asn	2252
CGT Arg	TTC Phe 435	GAC Asp	GGC Gly	AGC Ser	TTC Phe	GCT Ala 440	CGT Arg	GGC Gly	GTA Val	CCG Pro	TTC Phe 445	CAA Gln	TAC Tyr	AAC Asn	CCA Pro	2300
AGC Ser	ACA Thr 450	GGC Gly	GAC Asp	TGC Cys	CGT Arg	GTC Val 455	AGT Ser	GGT Gly	ACA Thr	GCC Ala 460	GCG Ala	GCA Ala	TTG Leu	GTC Val	GGC Gly	2348
TTG Leu 465	GCG Ala	CAA Gln	GAC Asp	GAT Asp	CCC Pro 470	CAC His	GCC Ala	GTT Val	GAC Asp 475	CGC Arg	ATC Ile	AAA Lys	CTC Leu	TTG Leu	TAC Tyr 480	2396
AGC Ser	ATT Ile	GCT Ala	TTG Leu	AGT Ser 485	ACC Thr	GGC Gly	GGT Gly	CTG Leu	CCG Pro 490	CTG Leu	ATT Ile	TAC Tyr	CTA Leu	GGC Gly 495	GAC Asp	2444
GAA Glu	GTG Val	GGT Gly 500	ACG Thr	CTC Leu	AAT Asn	GAC Asp	GAC Asp 505	GAC Asp	TGG Trp	TCG Ser	CAA Gln	GAC Asp 510	AGC Ser	AAT Asn	AAG Lys	2492

AGC GAC GAC AGC CGT TGG GCG CAC CGT CCG CGC TAC AAC GAA GCC CTG 2540  
 Ser Asp Asp Ser Arg Trp Ala His Arg Pro Arg Tyr Asn Glu Ala Leu  
 515 520 525

TAC GCG CAA CGC AAC GAT CCG TCG ACC GCA GCC GGG CAA ATC TAT CAG 2588  
 Tyr Ala Gln Arg Asn Asp Pro Ser Thr Ala Ala Gly Gln Ile Tyr Gln  
 530 535 540

GGC TTG CGC CAT ATG ATT GCC GTC CGC CAA AGC AAT CCG CGC TTC GAC 2636  
 Gly Leu Arg His Met Ile Ala Val Arg Gln Ser Asn Pro Arg Phe Asp  
 545 550 555 560

GGC GGC AGG CTG GTT ACA TTC AAC ACC AAC AAC AAG CAC ATC ATC GGC 2684  
 Gly Gly Arg Leu Val Thr Phe Asn Thr Asn Asn Lys His Ile Ile Gly  
 565 570 575

TAC ATC CGC AAC AAT GCG CTT TTG GCA TTC GGT AAC TTC AGC GAA TAT 2732  
 Tyr Ile Arg Asn Asn Ala Leu Leu Ala Phe Gly Asn Phe Ser Glu Tyr  
 580 585 590

CCG CAA ACC GTT ACC GCG CAT ACC CTG CAA GCC ATG CCC TTC AAG GCG 2780  
 Pro Gln Thr Val Thr Ala His Thr Leu Gln Ala Met Pro Phe Lys Ala  
 595 600 605

CAC GAC CTC ATC GGT GGC AAA ACT GTC AGC CTG AAT CAG GAT TTG ACG 2828  
 His Asp Leu Ile Gly Gly Lys Thr Val Ser Leu Asn Gln Asp Leu Thr  
 610 615 620

CTT CAG CCC TAT CAG GTC ATG TGG CTC GAA ATC GCC TGA CGCACGCTTC 2877  
 Leu Gln Pro Tyr Gln Val Met Trp Leu Glu Ile Ala \*

625 630 635

CCAAATGCCG TCTGAACCGT TTCAGACGGC ATTTGCG 2914

## (2) INFORMATION FOR SEQ ID NO: 2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 637 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Leu Thr Pro Thr Gln Gln Val Gly Leu Ile Leu Gln Tyr Leu Lys  
 1 5 10 15

Thr Arg Ile Leu Asp Ile Tyr Thr Pro Glu Gln Arg Ala Gly Ile Glu  
 20 25 30

Lys Ser Glu Asp Trp Arg Gln Phe Ser Arg Arg Met Asp Thr His Phe  
 35 40 45

Pro Lys Leu Met Asn Glu Leu Asp Ser Val Tyr Gly Asn Asn Glu Ala  
 50 55 60

Leu Leu Pro Met Leu Glu Met Leu Leu Ala Gln Ala Trp Gln Ser Tyr

[illegible]

His Asn Leu Pro Glu His Thr Ala Trp Val Asn Tyr Val Arg Ser His  
 385 390 395 400  
 Asp Asp Ile Gly Trp Thr Phe Ala Asp Glu Asp Ala Ala Tyr Leu Gly  
 405 410 415  
 Ile Ser Gly Tyr Asp His Arg Gln Phe Leu Asn Arg Phe Phe Val Asn  
 420 425 430  
 Arg Phe Asp Gly Ser Phe Ala Arg Gly Val Pro Phe Gln Tyr Asn Pro  
 435 440 445  
 Ser Thr Gly Asp Cys Arg Val Ser Gly Thr Ala Ala Ala Leu Val Gly  
 450 455 460  
 Leu Ala Gln Asp Asp Pro His Ala Val Asp Arg Ile Lys Leu Leu Tyr  
 465 470 475 480  
 Ser Ile Ala Leu Ser Thr Gly Gly Leu Pro Leu Ile Tyr Leu Gly Asp  
 485 490 495  
 Glu Val Gly Thr Leu Asn Asp Asp Asp Trp Ser Gln Asp Ser Asn Lys  
 500 505 510  
 Ser Asp Asp Ser Arg Trp Ala His Arg Pro Arg Tyr Asn Glu Ala Leu  
 515 520 525  
 Tyr Ala Gln Arg Asn Asp Pro Ser Thr Ala Ala Gly Gln Ile Tyr Gln  
 530 535 540  
 Gly Leu Arg His Met Ile Ala Val Arg Gln Ser Asn Pro Arg Phe Asp  
 545 550 555 560  
 Gly Gly Arg Leu Val Thr Phe Asn Thr Asn Asn Lys His Ile Ile Gly  
 565 570 575  
 Tyr Ile Arg Asn Asn Ala Leu Leu Ala Phe Gly Asn Phe Ser Glu Tyr  
 580 585 590  
 Pro Gln Thr Val Thr Ala His Thr Leu Gln Ala Met Pro Phe Lys Ala  
 595 600 605  
 His Asp Leu Ile Gly Gly Lys Thr Val Ser Leu Asn Gln Asp Leu Thr  
 610 615 620  
 Leu Gln Pro Tyr Gln Val Met Trp Leu Glu Ile Ala \*  
 625 630 635

## (2) INFORMATION FOR SEQ ID NO: 3:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "oligonucleotide"

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *Neisseria polysaccharea*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

CTCACCATGG GCATCTTGGA CATC

24

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "oligonucleotide"

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *Neisseria polysaccharea*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

CTGCCATGGT TCAGACGGCA TTTGG

25

Neisseria polysaccharea